②

RECURRENCE AND ERGODICITY FOR EXPONENTIAL
FAMILY STATE-SPACE MODELS

by

Gary Grunwald
Peter Guttorp
Adrian Raftery

AD-A213 466

DTIC
S **ELECTE**
OCT 17 1989
D ℃ **D**

TECHNICAL REPORT No. 173

August 1989

Department of Statistics, GN-22

University of Washington

Seattle, Washington 98195 USA

89 10 16 172

Recurrence and Ergodicity for Exponential Family State-Space Models

Gary Grunwald

University of Melbourne

Peter Guttorp
Adrian E. Raftery

University of Washington

ABSTRACT

We give two results concerning the properties of state-space models with exponential family observation distribution and conjugate state distribution. The first result gives a simple and general interpretation of the parameters of the predictive state distribution in terms of the observation forecast distribution. The second result shows how the first result can be used to check the long-term model properties of recurrence and ergodicity for a class of non-Gaussian observation distributions. In particular, these results apply to models with Poisson, binomial and multinomial observation distributions.

KEYWORDS: Bayesian forecasting; Binomial time series; Multinomial time series; Poisson time series; Recursive updating; Time series.

1. INTRODUCTION

The state-space approach provides a powerful formulation of many models for studying time series in the time domain. This approach encompasses Kalman Filter and ARMA models (Harrison and Stevens, 1976; Harvey, 1981, ch. 4) and has provided the standard approaches for treating missing values and irregularly spaced data (Jones, 1980 and 1981) and for computing the Gaussian likelihood for many models (Harvey and Philips, 1979). More

recently, a Bayesian view of state-space models has opened the way for modeling time series of non-Gaussian data. For example, useful models have been suggested for time series of binomial or Poisson observations (West, Harrison and Migon, 1985), and we have been exploring the case of compositional data (Grunwald, Raftery and Guttorp, 1989). These models provide a simple and computationally efficient framework for filtering and forecasting such series. In some cases they also allow the incorporation of covariates, trends, seasonality and interventions in a natural way.

In this paper we present results concerning the properties of some non-Gaussian state-space models. In particular, we study the recurrence and ergodicity of a class of models, and illustrate the results with the Poisson, binomial and multinonial examples. We also give a general interpretation of the parameters of the state distribution in terms of the observation forecast distribution.

2. NON-GAUSSIAN STATE-SPACE MODELING

A general state space model consists of three parts; an observation distribution, a state distribution and a method for making state forecasts. We assume an exponential family distribution for the observation and a state distribution that is conjugate to the observation distribution, but leave the state prediction rule unspecified for the present.

Let $y_t = (y_1^t, \ldots, y_d^t)^T$ be a vector in $\mathbf{R}^d$, and suppose that y_t follows an exponential family observation distribution with density

$$p(y_t | \theta_t) = \exp \{ y_t^T \theta_t - M(\theta_t) + S(y_t) \} I_{\mathbf{Y}}(y_t). \qquad (2.1)$$

Let Ψ be the interior of the convex hull of the support $\mathbf{Y}$ of the observation density, and assume that Ψ is a non-empty open set in $\mathbf{R}^d$. Let $\theta_t \in \Theta \equiv \{ \theta \in \mathbf{R}^d : M(\theta) < \infty \}$, the natural parameter space, where $M(\theta) \equiv \log \int \exp \{ y^T \theta \} p(y|\theta) dy$. Assume that Θ is a non-empty open set in $\mathbf{R}^d$. These conditions hold for the cases of most practical interest, where the observations have Poisson, binomial and multinomial distributions.

The conjugate density for the above observation distribution is

$$p(\theta_t \mid \sigma_{t\mid t}, \kappa_{t\mid t}) \propto \exp[\, \sigma_{t\mid t}\{\kappa_{t\mid t}^T \theta_t - M(\theta_t)\}\,]. \tag{2.2}$$

The subscripts and superscripts $t_1 \mid t_2$ for the parameters of this distribution describe the parameter at time t_1 given data and covariate information at all times up to and including t_2.

It follows from Theorem 1 of Diaconis and Ylvisaker (1979) that if $\sigma_{t\mid t} > 0$ and $\kappa_{t\mid t} \in \Psi$, then the density (2.2) can be normalized to a probability density and, for $\Theta = \mathbf{R}^d$, these are the only parameter values for which the state density is finite. When $\Theta \neq \mathbf{R}^d$ there may be other parameter values for which the density can be normalized. Usually in practice the observation density has some standard form and then the natural conjugate distribution is well known; for example, the beta is conjugate to the binomial, the gamma to the Poisson, and the normal to the normal when the variance is known.

To completely specify the state space model it remains only to describe a state prediction rule, or a method for obtaining the predictive state density $p(\theta_{t+1} \mid \sigma_{t+1\mid t}, \kappa_{t+1\mid t})$. These densities will act as the priors in the sequential applications of Bayes' Theorem to follow, and are specified through relations involving their parameters:

$$\sigma_{t\mid t} \rightarrow \sigma_{t+1\mid t}, \tag{2.3}$$

$$\kappa_{t\mid t} \rightarrow \kappa_{t+1\mid t}. \tag{2.4}$$

Such a rule may involve other parameters, covariate information or past data, and expresses what we "expect" to happen in the interval $(t, t+1)$, during which no new data are observed. This is where the temporal structure enters the model.

The model specification is now complete. Upon observing the new datum y_{t+1}, application of Bayes' Theorem yields the state posterior density, $p(\theta_{t+1} \mid \sigma_{t+1\mid t+1}, \kappa_{t+1\mid t+1})$, to be of the same form as (2.2) (the conjugacy property) with parameters

$$\sigma_{t+1\mid t+1} = \sigma_{t+1\mid t} + 1,$$

$$\kappa_{t+1\mid t+1} = \kappa_{t+1\mid t} + g_{t+1}(y_{t+1} - \kappa_{t+1\mid t}) = (1 - g_{t+1})\kappa_{t+1\mid t} + g_{t+1}y_{t+1}. \tag{2.5}$$

In (2.5), $g_{t+1} = 1/\sigma_{t+1|t+1}$ is analogous to the gain in the usual Kalman Filter.

The successive applications of Bayes' Theorem give recursions for the parameters of the posterior state distribution, and this procedure is known as filtering. The filtered distribution can often be thought of as the distribution of a signal that is to be observed with (possibly non-additive) noise. When both observation and state distributions are normal and the observation variance is assumed to be known, this procedure yields the usual Kalman filter. The predictive distribution of the next observation, or forecast density, is obtained by integrating out the state from the observation density (2.1), yielding

$$p(y_{t+1}|y^t) = \int_{\Theta} p(y_{t+1}|\theta_{t+1}) \, p(\theta_{t+1}|y^t) d\theta_{t+1}. \tag{2.6}$$

If the state prediction rule given by equations (2.3) and (2.4) is deterministic, then conditioning θ_{t+1} on y^t is equivalent to conditioning on $(\sigma_{t+1|t}, \kappa_{t+1|t})$. In some of the more common cases, this integration can be done easily and a simple closed form for the forecast distribution is then available (negative binomial for Poisson observations, beta-binomial for binomial observations and normal for Gaussian observations) while in other cases the integration may not be possible. In the following section we give a simple and general result for the mean of the forecast distribution in terms of the parameters of the state distribution. This result is used in proving the succeeding results concerning recurrence and ergodicity, and is also of interest in its own right as a point forecast of the observation.

The results we give concerning recurrence and ergodicity are based on Tweedie (1975). For a Markov chain $\{X_t\}$ on a normed space, he considers the quantity

$$\gamma_x \equiv E[\,|X_{t+1}| - |X_t|\, |\, X_t = x] = E[\,|X_{t+1}|\, |\, X_t = x] - |x|. \tag{2.7}$$

Intuitively, for "stable" processes, γ_x would be negative for x far from the center of the space, since then the future observation would be expected to be closer to the center than the present observation. In fact, he shows that, if $\{P(x, \cdot)\}$ is strongly continuous, the process $\{X_t\}$ is

Recurrent if and only if there is some $\alpha > 0$

with $\gamma_x \leq 0$ for all x such that $|x| > \alpha$. $\qquad$ (2.8)

Ergodic if and only if there is some $\alpha > 0$ such that $\gamma_x \leq -c$

for all x satisfying $|x| > \alpha$ for some positive constant c,

and γ_x is bounded above for all x satisfying $|x| \leq \alpha$. $\qquad$ (2.9)

In particular, when $|x| = |x_1| + \cdots + |x_d|$ is the L^1 norm and when the components of X_t are all either always positive or always negative, then γ_x is easily computed and direct evaluation of the long term properties of the model is possible. These ideas are the topic for the next section.

3. THE FORECAST MEAN, RECURRENCE AND ERGODICITY

In this section, we state and prove the two results of this paper. The first result gives a simple and general expression for the forecast mean. The second one uses this to compute the quantity γ_x of (2.7) in some cases, and thus allows a direct check of recurrence and ergodicity.

Theorem 1: Let $\Theta \subseteq \mathbf{R}^d$ be open and suppose that (2.2), (2.3) and (2.4) hold with $\sigma_{t+1|t} > 0$ and $\kappa_{t+1|t} \in \Psi$. Then

$$E[y_{t+1}|y^t] = \kappa_{t+1|t}, \qquad (3.1)$$

where $y^t \equiv (y_1, \ldots, y_t)$.

Proof: We compute the forecast mean directly and then apply Theorem 2 of Diaconis and Ylvisaker (1979).

$$E[y_{t+1}|y^t] = \int_Y y_{t+1} \, p(y_{t+1}|y^t) dy_{t+1}$$

$$= \int_Y y_{t+1} \left\{ \int_\Theta p(y_{t+1}|\theta_{t+1}, y^t) \, p(\theta_{t+1}|y^t) d\theta_{t+1} \right\} dy_{t+1}. \qquad (3.2)$$

By Fubini's theorem, whose application here is justified by the regularity of the exponential families and the finiteness of the integral in (3.2) as shown below, this is equal to

$$\int_{\Theta} \left\{ \int_{Y} y_{t+1} p(y_{t+1}|\theta_{t+1},y^t) dy_{t+1} \right\} p(\theta_{t+1}|y^t) d\theta_{t+1}$$

$$= \int_{\Theta} E[y_{t+1}|\theta_{t+1},y^t] \, p(\theta_{t+1}|y^t) d\theta_{t+1}$$

$$= \int_{\Theta} \nabla M(\theta_{t+1}) \, p(\theta_{t+1}|y^t) d\theta_{t+1}$$

$$= E[\nabla M(\theta_{t+1})|y^t], \tag{3.3}$$

where $\nabla M(\theta) = \left[\dfrac{\partial M(\theta)}{\partial \theta_1}, \ldots, \dfrac{\partial M(\theta)}{\partial \theta_d} \right]^T$. By Diaconis and Ylvisaker (1979, Theorem 2), the right-hand side of (3.3) is equal to $\kappa_{t+1|t}$. The equality

$$E[y_{t+1}|\theta_{t+1},y^t] = \nabla M(\theta_{t+1})$$

is standard exponential family theory (e.g. Bickel and Doksum, 1977, p.71) and is also given in equation (2.2i) of Diaconis and Ylvisaker (1979). □

Remark: One way to use the result (3.1) is to consider the difference $R_t \equiv y_{t+1} - \kappa_{t+1|t}$ as a residual at time t. This provides a starting point for model checking.

The second result of our paper concerns the long-term properties of some exponential family state space models. This result is a direct application of Theorem 1 and the results of Tweedie (1975), as reported in (2.7), (2.8) and (2.9) above, to the process $\{X_t\} \equiv \{\kappa_{t|t}\}$ of the parameters of the state density.

Theorem 2: Suppose that $\Theta \subseteq \mathbf{R}^d$ is open and that $\Psi \subseteq \mathbf{R}_+^d$ or $\Psi \subseteq \mathbf{R}_-^d$ where $\mathbf{R}_\pm^d = \{\mathbf{x} \in \mathbf{R}^d : \pm x_i \geq 0 \text{ for } i=1,\ldots,d\}$. Also, assume that (2.2), (2.3) and (2.4) hold where $\sigma_{t+1|t} > 0$ and $\kappa_{t+1|t} \in \Psi$ is a deterministic, time-invariant function of $\kappa_{t|t}$. Then

$$\gamma_k = s \, \mathbf{u}^T \left(\kappa_{t+1|t}(k) - \mathbf{k} \right), \tag{3.4}$$

and the long-term model properties are given by (2.8) and (2.9). In (3.4), $s = 1$ if $\Psi \subseteq \mathbf{R}_+^d$ and

$s = -1$ if $\Psi \subseteq \mathbf{R}_-^d$, and $\mathbf{u}^T$ is a d-vector of ones.

Proof: Using the L^1 norm, direct calculation of (2.7) for the Markov process $\{\kappa_{t\,|t}\}$ gives

$$\gamma_{\mathbf{k}} = E[\,|\kappa_{t+1\,|t+1}| - |\kappa_{t\,|t}|\,|\,\kappa_{t\,|t} = \mathbf{k}\,]$$

$$= s\,\mathbf{u}^T\,E\,[\{(1-g_{t+1})\kappa_{t+1\,|t} + g_{t+1}\mathbf{Y}_{t+1}\} - \kappa_{t\,|t}\,|\,\kappa_{t\,|t} = \mathbf{k}\,]$$

$$= s\,\mathbf{u}^T\,\{(1-g_{t+1})\kappa_{t+1\,|t}(\mathbf{k}) + g_{t+1}E[\mathbf{Y}_{t+1}\,|\,\kappa_{t\,|t} = \mathbf{k}] - \mathbf{k}\}\,. \qquad (3.5)$$

By the definition of the state-space model, $p(\theta_{t+1}|y^t) = p(\theta_{t+1}|\sigma_{t\,|t}, \kappa_{t\,|t})$. Thus (2.6) yields $p(y_{t+1}|y^t) = p(y_{t+1}|\sigma_{t\,|t}, \kappa_{t\,|t})$. It follows by Theorem 1 that the conditional expectation in the second term on the right-hand side of (3.5) is equal to $\kappa_{t+1\,|t}(\mathbf{k})$. We therefore obtain

$$\gamma_{\mathbf{k}} = s\,\mathbf{u}^T\,\{\kappa_{t+1\,|t}(\mathbf{k}) - \mathbf{k}\}\,. \qquad \square$$

Theorem 2 holds when the components of y_t are all positive or all negative and when the state prediction rule is deterministic and time-invariant. While these conditions may seem somewhat restrictive, there are many models whose long-term properties can be studied by these methods, as we illustrate with the examples below.

4. EXAMPLES

4.1 The steady model

The most common state prediction rule is $\kappa_{t+1\,|t} = \kappa_{t\,|t}$ and $\sigma_{t+1\,|t} = \delta\,\sigma_{t\,|t}$ with $0 < \delta < 1$, which Smith (1979) uses to define a "steady" model for non-Gaussian state space models. For this steady model, when the conditions of Theorem 2 apply, $\gamma_{\mathbf{k}} = 0$ so that any steady exponential family model of this type is recurrent, by (2.8). Such a model is ergodic, however, if and only if the space Ψ is bounded above in all dimensions, by (2.9). This last can be seen by taking, when Ψ is bounded, $\alpha = \max\{|x|:x \in \Psi\}$ so that (2.9) is satisfied trivially. In fact, these results show that any exponential family state space model satisfying the conditions of the theorems and having Ψ bounded in all dimensions is both ergodic and recurrent.

These results for exponential family state space models satisfying Theorem 2 are in contrast to the normal steady model defined when (2.1) and (2.2) are Gaussian. This normal steady model is neither ergodic nor recurrent, as can be seen either by noting that it is equivalent to a normal random walk (which has neither stability property) observed with Gaussian error, or by noting that there is an equivalent ARIMA(0,1,1) model, which is neither ergodic nor recurrent.

4.2 A Poisson time series model

The conjugate state density for the Poisson with mean λ_t specifies that λ_t has a gamma distribution with scale parameter $\sigma_{t|t}$ and shape parameter $\sigma_{t|t}\kappa_{t|t}$. The forecast distribution can be computed directly from (2.6) to be negative binomial with mean $\kappa_{t+1|t}$ (as in Theorem 1) and variance $\dfrac{\sigma_{t+1|t}+1}{\sigma_{t+1|t}}\kappa_{t+1|t}$. There does not appear to be a general result such as Theorem 1 for forecast variances.

The conditions of Theorem 2 apply (with $s = 1$) for appropriate state prediction rules. The steady model is recurrent but not ergodic by Section 4.1, since Ψ is unbounded. The theorem would also apply to other models specifying particular types of growth or decay. For example, for Poisson observations, $\kappa_{t+1|t} = \lambda\kappa_{t|t}$, along with the discounting $\sigma_{t+1|t} = \delta\sigma_{t|t}$ as before, specifies exponential decay $(0<\lambda<1)$, exponential growth $(1<\lambda)$ or the steady model $(\lambda = 1)$. Then $\gamma_k = (\lambda-1)k$, so that the process is not recurrent and hence not ergodic if $\lambda > 1$. Also, the process is ergodic and hence recurrent if $\lambda < 1$. For $\lambda \neq 1$, these models do not appear to have been studied in the literature, and Theorem 1 motivates them by giving an interpretation of the state prediction rules that define them. They may be useful for modeling the evolution of population size over time. They have the advantage of handling zero observations in a natural way, and they are flexible enough to be used together with other approaches such as threshold modeling. Care is needed in defining such models, since $\kappa_{t+1|t}$ must remain in Ψ.

4.3 Models for binomial and multinomial time series

We now consider the case where the observation distribution is multinomial, or binomial as a special case, with N independent of time. Then all of the conditions of the Theorems are satisfied (with $s = 1$) provided that the multinomial is expressed in a non-degenerate form in terms of the first d components for a $(d+1)$-component multinomial. The natural conjugate to the multinomial distribution is the Dirichlet distribution with parameters $\sigma_{t|t} \kappa_{t|t}^1, \cdots, \sigma_{t|t} \kappa_{t|t}^d, \sigma_{t|t}(n - \sum_{i=1}^{d} \kappa_{t|t}^i)$ and again the forecast distribution is a familiar one, the Dirichlet-multinomial (Mosimann, 1962). The forecast mean is given by Theorem 1, while the forecast variance of the jth component of the multinomial observation is $\frac{\sigma_{t+1|t}+1}{\sigma_{t+1|t}} \kappa_{t+1|t}^j (n - \kappa_{t+1|t}^j)$. As mentioned in Section 4.1, any multinomial state space model is both ergodic and recurrent.

References

Bickel, P. J. and Doksum, K. A. (1977) *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day.

Diaconis, P. and Ylvisaker, D. (1979) Conjugate priors for exponential families. *Ann. Statist.* 7 269-281.

Grunwald, G. K., Raftery, A. E. and Guttorp, P. (1989) Time series of continuous proportions. Technical Report no. 164, Department of Statistics, University of Washington.

Harrison, P. J. and Stephens, C. F. (1976) Bayesian forecasting. *J. R. Statist. Soc. B* 38 205-247.

Harvey, A. C. (1981) *Time Series Models*. New York: Halsted Press.

Harvey, A. C. and Phillips, G. D. A. (1979) Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika* 66 49-58.

Jones, R. H. (1980) Maximum likelihood of ARMA models to time series with missing observations. *Technometrics* **22** 389-395.

Jones, R. H. (1981) Fitting a continuous time autoregression to discrete data. In *Applied Time Series Analysis II* (D.F. Findlay, ed.), pp. 651-682. New York: Academic Press.

Mosimann, J. E. (1962) On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* **49** 65-82.

Smith, J. Q. (1979) A generalization of the Bayesian steady forecasting model. *J. R. Statist. Soc. B.* **41** 375-387.

Tweedie, R. L. (1975) Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stoch. Proc. Appl.* **3** 385-403.

West, M., Harrison, P. J. and Migon, H. S. (1985) Dynamic generalized linear models and Bayesian forecasting (with Discussion). *J. Amer. Statist. Assoc.* **80** 73-97.

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 173	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle) Recurrence and Ergodicity for Exponential Family State-Space Models	5. TYPE OF REPORT & PERIOD COVERED Technical Report 9-1-88 to 8-31-89
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s) Gary Gunwald Peter Guttorp Adrian Raftery	8. CONTRACT OR GRANT NUMBER(s) N-00014-88-K-0265

9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, GN-22 University of Washington Seattle, WA 98195	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N 63374 1107 NE 45th Street Seattle, WA 98105	12. REPORT DATE August 1989
	13. NUMBER OF PAGES 12

14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, If different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and Identify by block number)

Bayesian forecasting; Binomial time series; Multinomial time series;
Poisson time series; Recursive Updating; Time series.

20. ABSTRACT (Continue on reverse side If necessary and Identify by block number)

We give two results concerning the properties of state-space
models with exponential family observation distribution and conjugate
state distribution. The first result gives a simple and general inter-
pretation of the parameters of the predictive state distribution in terms
of the observation forecast distribution. The second result shows how
the first result can be used to check the long-term model properties
of recurrence and ergodicity for a class of non-Gaussian observation

DD FORM 1 JAN 73 1473   EDITION OF 1 NOV 65 IS OBSOLETE   (con't on next page)
S/N 0102-LF-014-6601

distributions. In particular, these results apply to models with Poisson, binomial and multinomial observation distributions.